Identifying businesses and entrepreneurs from the Censuses, 1891-1911

Harry Smith, Robert J. Bennett and Carry van Lieshout

hjs57@cam.ac.uk rjb7@cam.ac.uk cv313@cam.ac.uk

Working Paper 4:
Working paper series from ESRC project ES/M010953:
Drivers of Entrepreneurship and Small Business

University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure, Downing Place, Cambridge, CB2 3EN, UK.

May 2017

Comments are welcomed on this paper: please contact the authors by email as above.

Keywords: Entrepreneurship, Employers, Self-employment, Small businesses, Census

JEL Codes: L26, L25, D13, D22

Identifying businesses and entrepreneurs from the Censuses, 1891-1911

Harry Smith, Robert J. Bennett and Carry van Lieshout

Working Paper 4: ESRC project ES/M010953: Drivers of Entrepreneurship and Small Businesses, University of Cambridge.

1. Introduction

This paper explains how employers and own account self-employed individuals can be extracted from the original manuscript records of household returns to the population census for 1891-1911. The database for Entrepreneurs 1851-1911 referred to in this and other project Working Papers for ESRC project ES/M010953 *Drivers of Entrepreneurship and Small Businesses*, is an amalgamation of several sources. The data referred to in this working paper for 1891-1911 is derived from the Integrated Census Microdata (I-CeM) deposited at UK Data Archive (UKDA).[1] These records are derived from the transcriptions made by the commercial genealogy provider Find My Past (part of BrightSolid) in conjunction with The National Archives (TNA). This paper describes how individual entrepreneurs can be identified and extracted from the original Census Enumerators Books (CEBs). The population census was not a business census. It was designed by the General Register Office (GRO) to count the population. As a result, the way in which the census gathered material constrains the sort of employer information that can be obtained. The material collected in each census and its value for the identification of entrepreneurs as a raw data base and published tables are described in WP 2 from the ESRC project: *Employers and the self-employed in the censuses 1851-1911: The census as a source for identifying entrepreneurs, business numbers and size distribution*. The method of extraction for the censuses 1851-81 are described in WP 3. An overview of the project is given in WP 1: *Drivers of Entrepreneurship and Small Businesses: Project overview and database documentation*.

[1] Schürer et al, 2016; Higgs et al, 2015.

The 1891-1911 censuses are significantly different in structure from those carried out between 1851 and 1881. They specifically identified employers and the self-employed. However, they did not contain information about employee numbers or acreage. This paper describes the how entrepreneurs are identified in each census, the issues that arise, and the method by which each census was cleaned and prepared for subsequent analysis.

2. The Census Form and Report

2.1 The 1891 Census

1891 represents a break in census-taking practice from the point of view of identifying business owners and the self-employed. It was argued by the 1891 Census compilers that the old question about employers (which required employers to identify themselves by their occupational descriptor) was 'rarely' answered. However, as analysis of the earlier censuses shows, this was not true. The reluctance to deal with this question seems to have arisen partly due to the cost of analysing the data (an analysis which had not been carried out since 1851). However, as discussed in WP 2, it also reflected different understandings of the purpose of the Census. The GRO continued to see the census as a primarily demographic exercise, and hence, concerned with the collection of socio-medical information. However, interests outside the GRO increasingly wanted the Census to address economic issues. Thus, following some of the recommendations of a Treasury Committee enquiry into the taking of the Census, a number of changes were made to the Census form. Most important for this discussion was the inclusion of a question about employment status.

The new question asked householders to report all employed residents' employment status by putting a cross in one of three columns (numbered 7, 8 and 9) headed: 'employers', 'employed', or 'neither employer or employed'.[2] These three columns were grouped with the occupation descriptor under a heading 'Professions or Occupation'. This should have resulted in all sectors of employers and self-employed replying. Since it also applied to all residents, it should have covered both genders and all ages. However, in the general instructions to householders it was stated that 'These three columns 7, 8, and 9, refer only to employment in

[2] Schürer, 1991, 20-26.

trades and industries, and not to the employment of servants.'[3] This is the most explicit exclusion of other sectors for the period, and might suggest that the 1891 Census would give little information about non-trade and non-industrial sectors. However, analysis of the data provided by the census enumerator books (CEBs) reveals that most people in non-trade and non-industrial occupations returned employment status information. The data were tabulated and published for most trades and industry, but not tabulated for other categories by the Census Office in the 1891 Report. However, information about employers in the professions, mining, transport and other sectors remained in the underlying data.

The rest of the specific information contained in the general instructions and relating to employment status was as follows:

> A *cross* must be made in Column 7, headed 'Employer', when a person is a master, employing under him workers in his trade or industry; in Column 8, headed 'Employed', when the person is working in a trade or industry under a master; and in Column 9, headed 'Neither Employer nor Employed', when the person neither employs other workmen in his trade or industry, not works for a master, but works on his own account. Married women assisting their husbands in their trade or industry are to be returned as 'Employed'.[4]

This clearly set out the distinction between the three kinds of employment status and also explicitly included married women; however, they may have been undercounted depending on how householders interpreted the instruction 'assisting their husbands'. The main limitation, compared to the pre-1891 Censuses, is the absence of information on employee numbers, and on acreages for farmers.

The 1891 Census Report was critical of the quality of information provided by the new question about employment status. It argued that in many cases no cross was entered in any column. Where the columns were marked, the Report reported that two or three of the columns were sometimes crossed or that the wrong column (in the view of the GRO) was ticked. The Census Report also suggested that this issue arose from 'the foolish but very common desire of persons to magnify the importance of their occupational condition.' This resulted in the 'the otherwise unintelligible fact' that some occupations had more 'employers than employed, more masters than men'. The Report particularly highlighted 'Builders,

[3] 'General Instructions', The National Archives (TNA), RG 27/6, Census of England and Wales, Householder's Schedule, 1891.
[4] Ibid.

Provision Dealers, Coal Dealers, Road Contractors, Dealers in Hemp, &c., Dealers in Cane, Rush, &c. and others.' The Report did not just attribute the supposed failings of the returns on employment status to falsehood and embellishment; it also suggested that a lack of familiarity with filling forms and general inability to cope with complex instructions would have led to many genuine mistakes.[5]

This decidedly downbeat evaluation of the quality of the 1891 Census information regarding employment status is unduly pessimistic. As will be shown below, the rate of double ticking was very low and the ratios of employers, the self-employed and the workers seem reasonable, both at an aggregate level and broken down by occupation, contrary to the Census Report's statements. Furthermore, as Kevin Schürer calculated from the Report tables, and the I-CeM data now demonstrates, the extent to which people failed to give an answer to the employment status question was exaggerated in the Report itself.[6]

The 'Special Instructions' to householders were changed slightly in 1891. Unlike in previous censuses, there were no special instructions for those in 'trade and manufactures' and farming; instead, they were meant to respond by following the other instruction, e.g. to identify their specific branch of activity. The only mention of farming came in instructions 9 and 10 which dealt with, respectively, how to return sons and other relatives of farmers, and how to treat agricultural labourers. These differences may have led to different types of responses from those engaged in 'trades and manufactures' and farming compared to previous censuses.

With regards to the 'General Instructions', for those individuals with multiple occupations, the instruction (2) in 1891 was similar to that given in 1871 and 1881: 'A person following several distinct occupations must state each of them in the order of their importance'. The instruction concerning women and children (4) also retained a similar wording to that given in 1871 and 1881. However, there was no specific instruction on how partners should respond, which must have resulted in a number of multiple entries for a given firm.

The results of the 1891 question were published in summary form; the first time an employment status summary table was published since 1851. The published table broke the

[5] *Census of England and Wales, 1891, Vol. IV General Report, with Summary Tables and Appendices*, *Parliamentary Papers*, 1893-4 (CVI), 36.
[6] Schürer, 1991, 26.

data down by gender. However, for a number of occupation classes no data on employment status was given. Thus, there were no entries for professions (Class I), domestic (Class II), and merchants and agents, dealers in money, and insurance (Class III, Order 5, Sub-Orders 1-3). Some orders were only partially tabulated, thus railway and water-borne conveyance (Class III, Order 6, Sub-Orders 1 and 3), messengers and porters (Class III Order 6, Sub-Order 5), agriculture and fishing (Class IV), some weavers (Class V, Order 17, Sub-Order 5), mining (Class V, Order 21, Sub-Order 1) and mechanics and labourers (Class V, Order 22, Sub-Order 2) were missing from otherwise complete data. The I-CeM data allows us to add this missing information for the first time.

There were some shifts in the occupation classification between 1881 and 1891. In general these changes reduced the overall number of occupational categories by amalgamating headings within sub-orders. For example, civil engineers and mining engineers were combined to create the category 'Civil and Mining Engineers'.[7] This reduced the number of occupational categories from 399 to 347.

2.2 The 1901 Census

Despite the complaints about the employment status question in the 1891 Census Report, the question was retained in 1901. However, the form of the question was changed. Instead of ticking one of three columns, householders now had to fill in a single column in which they instructed to:

> Write opposite the name of each person engaged in any trade or industry, either
> (1) '*Employer*' (that is, employing persons other than domestic servants)
> (2) '*Worker*' (that is, a worker for an employer), or
> (3) '*Own Account*' (that is, neither Employer nor working for Employer, but working on own account).[8]

In addition, following requests from the Home Office and the Board of Trade, a new question was added which asked individuals to state whether they carried on their occupation in their

[7] *1891 General Report*, 133-5.
[8] *Census of England and Wales, 1901. General Report, with Appendices*, *Parliamentary Papers*, 1904 (CVIII), 322.

own home.[9] The GRO's continued lack of interest in employment status was manifest in the fact that the Census Report contains no comment on the information provided by these questions and no comment on the quality of the returns. The 21 general and special instructions remained much the same as they had been in 1891, there continued to be no specific discussion of farming.

The results of the 1901 question were published in summary for men and women as in 1891, broken down by workers, employers and own account. In addition, these were also broken down by working at home and not at home. The same occupational categories were excluded from the published tables as in 1891. Again, this represents a conscious decision by the GRO not to tabulate some occupations, and we can now fill in these blanks using the I-CeM data.

There were further changes to the occupation classification. Again, discussions with the Home Office and Board of Trade drove these changes, and the number of categories increased from 347 in 1891 to 382 in 1901.[10] This was not just a process of splitting previously amalgamated categories, it also saw some small categories removed and other subdivided, for example, 'Persons engaged in Iron Manufacture' were divided into those employed in 'Blast Furnaces', 'Puddling Furnaces and Rolling Mills', 'Steel Smelting and Founding', 'Iron Founding' and in the production of specific iron articles. The Census Report also claimed that an attempt was made to separate makers and dealers; however, the effort was limited in the 1901 census and not entirely successful.[11]

2.3 The 1911 Census

The 1911 Census is the first Census where the original householder schedules are available rather than the Census enumerator books, which provide the data for all previous censuses. This results in wider variation in quality and style of responses since the returns were not screened and aligned with the census instructions by enumerators as in the CEBs, although clerical marks do appear. The variations reflect individual householders' perceptions of how they wished to describe their occupations, which offer potentially valuable insights not available in the CEBs, but lead to a much wider range of possible answers. The questions with

[9] Ibid., 74, 322.
[10] Ibid., 73.
[11] Ibid., 74.

regard to employment status and at home remained the same as in 1901. It was the first census to ask for the business or public body which employed each individual. This was collected to allow a more accurate industrial classification to be created.[12] This marked a break with all previous census processes and the beginning of the foundation of the format for the modern censuses' question about employment status.

There were changes in the instructions to householders. The multiple occupation question changed substantially. Instead of asking for all occupations to be listed, the instruction now stated 'If more than one Occupation is followed, state that by which living is mainly earned.' Additionally, the instructions to householders provided explicit directions to record the occupations of women 'engaged in any business or profession'. In this regard, 1911 reflected a return to the structure of instructions given before 1891. Specific instructions about farmers also made a return, with householders instructed to state whether '"Farmer," "Grazier," or "Farm Bailiff"', and additional instructions were given for describing relatives of farmers engaged in farm work.[13]

The 1911 Census Report, as with the 1891 and 1901 Reports, tabulated the employment status data according to gender and also recreated the 'at home' breakdown similar to 1901. Once again the same occupation categories were not tabulated with a few notable exceptions. Farmers (Class VII, Order 1, Sub-Order 1), washing and bathing services (Class IV, Order 3, Sub-Order 8) were both broken down by employment status for the first time.[14] Also for the first time, there was a partial published regional breakdown by employment status for London, Lancashire and Yorkshire, but not for any other counties.

The occupational classification changed once again in 1911. The number of categories increased from 382 in 1901 to 473 in 1911. This considerable increase in categories resulted from the move to machine-tabulation of the census data. This development allowed the Census to ask a wider range of questions and to produce more detailed tabulation and analysis.[15]

[12] *Census of England and Wales, 1911. General Report with Appendices*, *Parliamentary Papers*, 1917-18 (XXXV), 97-8.
[13] Ibid., 257.
[14] Ibid., 14-17.
[15] Ibid., 99; Higgs, 1996, 415-20.

3. Checking and Cleaning the Census

The later censuses require far less processing than the 1851-81 censuses because the employers and own account individuals are already identified by the employment status question. However, checks are still required to evaluate the reliability of the data, and cleaning is still required to remove spurious employers and to ensure the data can be compared with other years.

3.1 The 1891 Census

The primary check required for 1891, given the GRO's negative comments, is to evaluate the extent to which the self-reporting of employment status was flawed. This check has three aspects. First, to check the number who provided no answer to the employment status question. Secondly, to ascertain how common it was for individuals to tick more than one column when answering the employment status question. Thirdly, evaluating the Census Report's claim that people exaggerated their employment status.

Non-response to employment status

There are 29,050,639 people in the 1891 I-CeM Census database. This compares with 29,002,525 in the published GRO report. In I-CeM, 10,573,570 people returned no occupation; 7,476,315 of the remaining 18,477,069 gave no answer to the employment question. This number includes many individuals who returned an answer to the occupation question, but whose answers were not economically active occupations. For example, it includes more than 4.6 million scholars and over 450,000 people who were living on their own means. Removing those for whom not answering the question was legitimate leaves around 1.5 million people with occupations who did not give their employment status; 13 per cent of the 11.8 million individuals with genuine occupations. There were also people who had an answer to the status question but no occupational descriptor (see discussion below).

Organising the I-CeM data by the 1891 occupation classification and calculating the proportion of blank answers to the employment status question allows the mean percentage of blank answers to be calculated. Without removing any occupations the mean is 12.07 per cent. However, the largest proportion of blank returns to the employment status question are

found in those occupation categories in which we might expect legitimate blank answers (those supported by private means, the retired, students and so on). If we remove these categories with very high proportions of blank returns, then the mean drops from 12.07% to 9.01%. Thus, a small, but not insignificant proportion of the employed population failed to provide an answer to the employment status question. The remaining blanks were predominantly younger (under 40, and especially under 30), older and female. The pattern to a large extent tends to reflect relationship of those resident in a household with another person as head: i.e. they were younger adult sons and daughters or older relatives. This must be born in mind when considering the statistics derived from the 1891 I-CeM data. However, the proportion is not as high as implied by the comments in the GRO 1891 Census Report.

It might be argued that failure to correctly answer this question was correlated with lower social status and that consequently it would over-count employers. However, as shown in Table 1, the rank frequency of the age profile and gender proportions of those who provided no answer to this question does not match the age profile of any of the three employment categories, suggesting they is a random mix of employers, own account and workers with blanks. Statistical tests on Table 1, with the null hypothesis that there is no relationship of the ranked frequency of blanks to each of the response categories for employment status by age, demonstrate that the null hypothesis is generally accepted: that non-responses do not have the same pattern by age as employers or own account, although the null hypothesis is closer to being accepted for workers, the total occupied, and for females. Table 1 reports the Spearman ranked correlation by age of two-way comparisons of blanks against each category (reported at the foot of each column for each gender), and blanks against all occupied which is the sum of all categories (at the foot of the 'blank' column). The only significant relationship is between the proportion of all female occupied and blanks, although the relationship between both male and female workers and blanks is also significant at the lower level of p=0.1. This indicates that female blanks are like the pattern of workers and all occupied in terms of rank frequency by age. However, generally the blank responses are randomly distributed by age and gender compared to each entrepreneur category; if there is a relationship it is closest to the distribution of workers and the all occupied. Indeed, the 1891 Census Report itself argued that the propensity to fill in forms incorrectly was no respecter of education or social status.[16] Further statistical tests on the blanks are reported below, and in subsequent publications.

[16] *1891 General Report*, 36.

Age	*Male*				*Female*			
categories	*Emp*	*OA*	*worker*	*blank*	*Emp*	*OA*	*worker*	*blank*
15-19	2934	10475	1129256	112668	1649	24696	530151	111321
20-29	71817	81133	1869007	182547	11670	80586	643092	192941
30-39	151243	115161	1363718	126820	16797	65705	266748	102594
40-49	154547	114847	954593	96489	20628	66611	175215	81578
50-59	121503	98422	598155	73935	20606	59544	113915	68933
60+	102818	108834	416889	126345	22023	57445	82969	110971
Spearman	-0.34	0.03	0.63	0.60	0.54	0.03	0.66	0.83*

Table 1. Non-response (blanks) to employment status question 1891 compared to age for employers, own account, workers and all occupied; 2-way ranked Spearman correlation significance tests (* indicates significant at p=0.05).

An additional issue is whether there was any geographical pattern of non-responses. The geography of blank answers to the employment status question suggests a few concentrations at such high levels that they can only be the result of enumerator failings rather than individuals failing to answer the question. Figure 1 shows that there were particular groups of registration sub districts (RSDs) where there were high proportions of blank answers; west Wales is the worst affected area, particularly the counties of Cardiganshire, Carmarthenshire and Pembrokeshire. When examined at a finer geographical level it is apparent that there are enumeration districts in which few individuals are given an employment status. For example, the parish of Llanbadarn-Odwyn in Cardiganshire is covered by a single enumeration district. Of the 260 people recorded in Llanbadarn-Odwyn in the 1891 Census, only seven have an employment status.[17] It is unclear why these seven were given employment statuses and other were not. Indeed, one who is given a status, David Evans a farm servant who is returned as a worker, is directly below an individual with the same occupation for whom no employment status is given. Perhaps the enumerator, one David Morgan, did not think it an important question and consequently did not press those who provided no answer to provide one. It should be noted that Morgan, who was a farmer, himself did not return an answer. In this parish, and the others that made up the Llangeitho RSD the quality of the enumerators did not improve in 1901, when once again very few answers to the employment status question were recorded. In part this was because many of the same enumerators carried out the 1901 Census.

[17] Manuscript CEB, 1891 Census of England and Wales, TNA, RG12/4563/1 – RG12/4563/51/9.

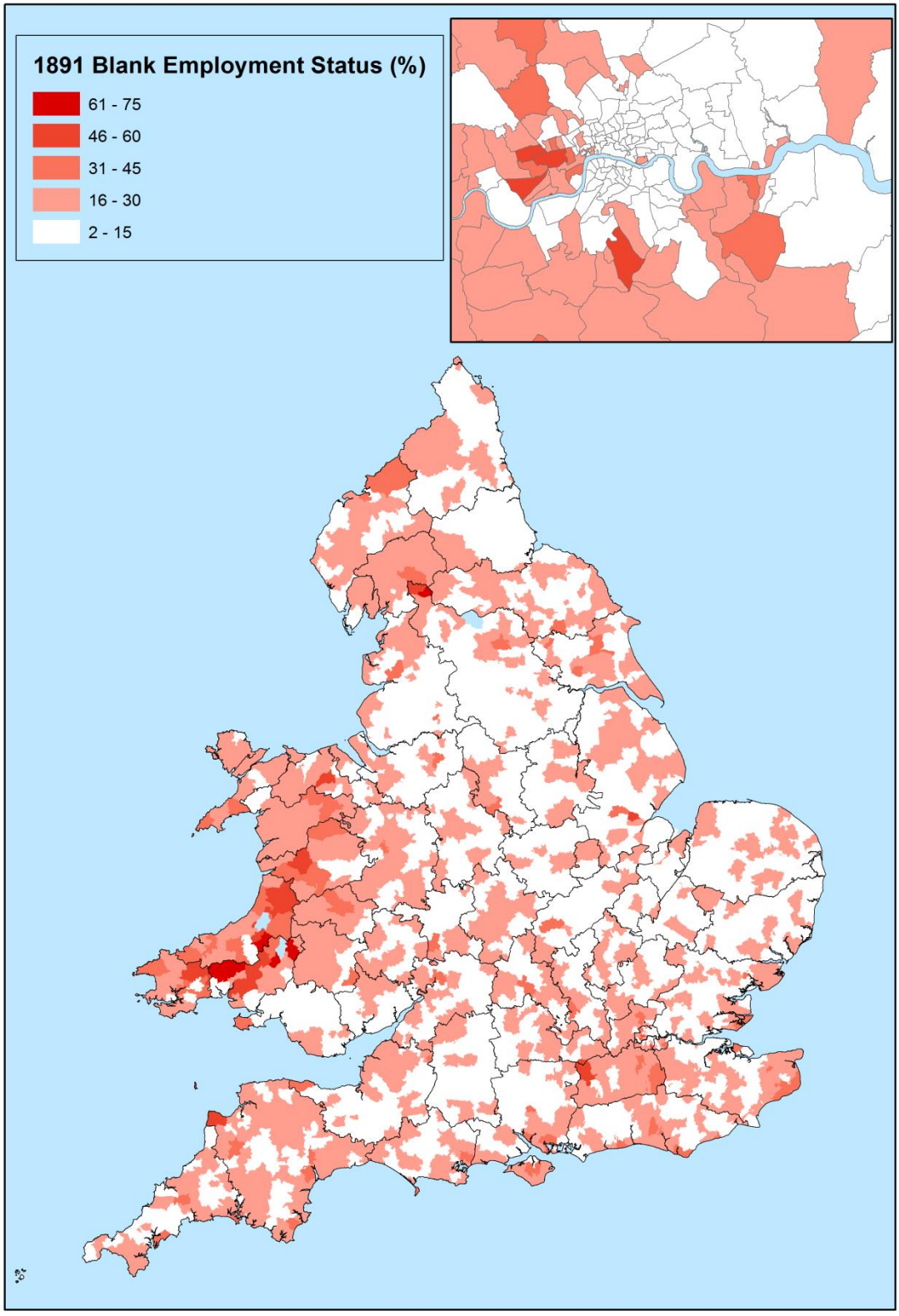

Figure 1. Blank employment status answers, percentage by RSD, 1891.

Note: The following occupation descriptors removed: '-', '*servant*', '*service*', '*retire*', '*pension*', '*own mean*', '*student*', '*scholar*'.

The variations in RSD enumeration rates also support the suggestion that the tendency not to answer the employment question was spread randomly throughout the population. RSDs with few employers tend to also have few workers and own account individuals. For example, the RSD with the lowest enumeration rate was Ramsgill in the West Riding of Yorkshire. This RSD had a very small number of people who provided an answer to the employment status question. Just 48 people had an employment status (out of 285 occupied in total). However, these were spread between each employment category with 29 workers, 8 employers and 11 own account. Although this is in the likely rank order of frequency, and has approximate proportions of the population that are plausible, it appears to significantly understate the proportion of workers. However, there are some RSDs where the numbers might suggest that employers were under-recorded. One such RSD is St John in St George in the East (London) where there were 2,951 workers and 133 own account, but only 28 employers. However, this fits with the nature of this particular area of London where most of the residents were engaged in casual or irregular labour, or artisanal work in the furniture and clothing trades. Indeed, Charles Booth found that employers made up only 3.13 per cent of St George in the East's population.[18] St George North, the other RSD which constituted St George in the East, had a much higher enumeration rate for the status question (88 per cent), but shows a similar balance between employers, own account and workers. There is one RSD which does look anomalous when considering the over or under-enumeration of employers or own account individuals, Llangeitho, where the poor quality of enumerators has already been noted. In this RSD there were more people returned as own account than were returned as workers: 74 workers as opposed to 106 own account.

There can be considerable variation within RSDs. For example, in the RSD of Saffron Hill 93.5 per cent of those with occupations were given an employment status. However, this high level of enumeration masks significant variation in the quality of enumeration in each of the component parishes. Two of the parishes within the RSD had similarly high levels of enumeration: Hatton Garden (96.3 per cent) and St Sepulchre (93.8 per cent). The third parish, Charterhouse, however, had a very low enumeration rate; it was just 1.5 per cent. This is because Charterhouse was not a real parish, but instead the return for the school. However, other RSDs have similarly significant variations which cannot be explained by such unusual parish structures. Thus, the RSD of Bingham in Nottinghamshire has an overall enumeration rate of 81.4 per cent. The constituent parishes, however, range from 100 per cent (the parishes

[18] Booth, 1892-1903, i, 64-5, 80-81.

of Shelton and Wiverton Hall) to 2.5 per cent (Kinoulton). The majority of RSDs do not contain such dramatic variation. However, the data clearly reveal differences in the likely extent of enumerator variation across the country.

The issue of enumerator variations is assessed further in another Working Paper which compares I-CeM, CEB entries and local directory entries in a series of case study areas. This assessment shows the tendency for some enumerators to be less thorough than others in recording status, with many making implicit assumptions about common occupations, or lack of them. This results in wide variations between small areas. However, as in the above analysis, there is little if any systematic bias in almost all the cases examined. The exceptions are in a few possible cases where it is clear the enumerator did not understand the question and recorded high numbers in occupations that must be entirely incorrect. In the sample of 63 enumeration districts, one enumerator was detected as completely arbitrary and inaccurate, e.g. assigning employer status to domestic servants and farm labourers but not to shopkeepers and innkeepers who were generally employers or own account. More common was a more generalised omission of status responses (blank responses) for the whole enumeration district, or almost all within it. This further confirms that there is no systematic bias to the general incidence of non-responses, but there will be a few enumeration districts and perhaps a larger part of a few RSDs where responses are very unreliable. This has to be allowed for in subsequent analysis. The blank answer problem means that the census under-records some entrepreneurs as well as workers.

Multiple 'ticks'

If the issue of blank answers is a real but manageable issue, the matter of double ticks is not a significant problem. Checks were carried out by consulting CEBs in London, Brighton, Birmingham and Oldham and searching for individuals who had more than one tick in the employment status columns. Approximately 36,156 individuals were checked and only 33 double ticks were found (0.09 per cent). The 33 double ticks were treated in the following ways by FindMyPast in I-CeM:

Error	*N*	*Error rate (%)*
Coded first of two ticks	10	0.027
Coded second of two ticks	5	0.013
Two ticks coded as blank	3	0.008
Clerks' changes followed	12	0.033
Clerks' changes ignored	3	0.008
Overall total	33	0.091

Table 2. Double Ticking in 1891; I-CeM compared to CEBs.

Similarly, the Working Paper comparing I-CeM with CEBs and local directory entries shows that 14 individuals out of 2817 employers and own account had double ticks (0.5%), which is higher than the comparisons above, but still much lower than claimed by GRO. Moreover all of these entries were entirely reasonable ways of recording the individual's multiple statuses given their census statement, and/or directory information. For example: 'solicitor & clerk to boards', 'fishmonger & grocers traveller', and 'farmer coal merchant manager' which were all employer as well as worker statuses; or 'weaver clothier & hair dresser' one of which was own account and the other a worker; and 'innkeeper & tinman (& coasting pilot)' which appears to be one own account and either a further own account and a worker activity, or two further worker activities. Indeed, the number of multiple ticks appears to be far lower than the probable number of multiple occupations likely in the population. Hence, whilst the procedures followed by FindMyPast and I-CeM in coding individuals with two ticks are rather random, the number of double ticks appears to be very small but meaningful, contrary to the claims of the GRO 1891 Census Report.

Embellishment of employment status

It is difficult to test directly the claim that individuals embellished their employment status. However, we can consider the ratios of employers, own account and workers to see whether they seem plausible compared with other sources. In 1891 the numbers of employers, own account and workers over the age of 15 years was as shown in Table 3.

Status	*N*	*%*
Employer	721,114	6.95
Own Account	946,108	9.11
Worker	8,709,763	83.93

Table 3. Employment Status of over 15s, 1891 (I-CeM).

These proportions seem broadly plausible in rank order of frequency. However, the number of employers is high compared with modern data. This impression is confirmed when these figures are compared to 1881 and 1901. The percentage figure for employers in 1891 is higher than in both 1881 and 1901. Furthermore, the actual number of employers is higher in 1891 than in 1901: 721,114 in 1891 compared to 587,332 in 1901. This decrease in the number of employers is accompanied by an increase in the number of own account individuals (from 946,108 to 1,213,611). Part of these shifts will have arisen from real changes in the structure of the economy, but it seems likely that the high number and proportion of employers in 1891 was partly due to people either preferring to tick employer than own account, possibly to inflate their economic status, or because the column headings to the three columns for employer, worker and own account were perhaps somewhat ambiguous.

The argument made by the Census report to the effect that some occupations had 'more employers than employed' seems not to be supported. This appears to be a wilful mis-understanding of the occupational categories. For example, builders do have more employers than workers (17,609 employers compared to 16,415 workers). However, this is unsurprising given that the majority of workers involved in the building trade were returned under their specific trade, carpenters, masons, bricklayers and so on. All of those occupation categories showed far greater numbers of workers than employers or self-employed people; for example, there were 176,446 workers in the carpenter category as opposed to 12,203 employers. Other occupations cited in this category included provision dealers, where the majority are own account individuals, again unsurprising as provision dealers were mainly people running their own shops. Indeed, dealers of all kinds tended to have high proportions of employers and own account, something which is logical when the structure of those trades is considered. Again, the Census Report comments seem to reflect the GRO's reluctance to include this question more than any major problem with the data.

Comparison of I-CeM and published data

It was also valuable to compare to the aggregate figures produced by the I-CeM data with the equivalent data provided in the 1891 Census Report (see Tables 4 and 5).

	I-CeM	*1891 Report*	*Difference*
All	21,927,624	21,969,578	-41,954
Employers	463,525	500,189	-36,664
Own Account	656,667	672,807	-16,140

Table 4. Employment status of over 10s, 1891 (I-CeM).

Note: The I-CeM employer and own account counts are made of those occupations which were tabulated by employment status in the 1891 Census, and excludes professionals, farmers and so on.

	I-CeM	*1891 Report*	*Difference*
Male	10,511,981	10,591,967	-79,986
Female	11,435,887	114,61,890	-26,003
Unknown	150,743	n/a	

Table 5. Gender of over 10s, 1891 (I-CeM).

In both cases the differences between the I-CeM figures and the Census Report are negligible when considered at the national level. The comparison of employer/own account ratios is taken further in a Working Paper on long-term trends.

3.2 The 1901 Census.

The employment status question changed in 1901; householders had to write whether they were an employer/own account/worker, instead of ticking a column. Consequently, the issue of double ticking is not important for this census. However, it is still necessary to judge whether the question was consistently and correctly answered.

Once again, a significant number of people did not answer the question. 20,444,043 people out of the total population of 32,493,071 provided no information about their employment status. 16,477,173 of these people gave no occupation, leaving 3,966,870 individuals who did report an occupation but have no employment status. As in 1891, most of these individuals reported occupations which were not economically active occupations: students, the retired and those living on independent means. Removing those leaves 2.3 million people with real occupations but no employment status, 16 per cent of the 13.9 million economically active. The frequency distribution of the age profile of those with blank employment status again does not match the profiles of either workers, employers or own account individuals.

Age	*Male*				*Female*			
categories	*Emp*	*OA*	*worker*	*blank*	*Emp*	*OA*	*worker*	*blank*
15-19	3147	10783	1244140	198629	1573	25235	617698	194333
20-29	58176	107223	2210846	308322	7063	97411	737899	346259
30-39	133604	172612	1590302	206902	10765	89657	271079	158675
40-49	134652	164687	1131411	146343	12454	88409	171635	103982
50-59	104660	139747	705446	98775	12656	77810	104051	76794
60+	81889	136910	459214	133977	14412	75371	77631	114079
Spearman	0.26	-0.14	0.94*	1.0*	0.77	0.68	0.77	1.0*

Table 6. Non-response (blanks) to employment status question 1901 compared to age for employers, own account, workers and all occupied; 2-way ranked Spearman correlation significance tests (* indicates significant at p=0.05).

Table 6 tests the comparison statistically following the same pairwise Spearman rank correlation tests as in Table 1. In this case the null hypothesis is accepted that blanks are again randomly distributed by ranked age compared to the distribution of employers and own account at p=0.05. However, the null hypothesis for male workers, all occupied males, and for all occupied females suggests that blanks generally reflect the frequency pattern of the workers and total occupied population. At the lower significance level of p=0.1 all-female blank categories reject the null hypothesis. Whilst the general results indicate that there is no significant pattern related to employers and own account proportions by age and gender, in 1901, these results indicate that as in 1891 there is some tendency to use blanks instead of

indicating workers. Most worrying in 1901, however, is a tendency for blanks to be used systematically for female workers and own account. This suggests that there may have been some prejudice by household heads or enumerators towards using blanks for worker statuses rather than acknowledging female employment (which was often in the home; see further below). Hence, although generally removal of the blank employment status individuals should not bias interpretation for entrepreneurs when working at the 95% level of significance, we must bear in mind a potential bias against recognising female entrepreneurial status and general under-recording of female occupations.

The geography of those with blanks for this question is more widespread and less concentrated than in 1891. Many RSDs that were in the first category (0-18 per cent blank answers) in 1891, now appear in the second category (18-35 per cent). The more dispersed pattern suggests that under-enumeration is less of a problem in 1901 than it was in 1891. Widespread low levels of blank answers suggests people were failing to answer the question rather than enumerator error. Thus, the 3.9 million missing answers are more likely people who misunderstood the question or refused to answer for one reason or another. However, enumerator error is unlikely to have disappeared completely and the similar concentrations of blank answers in West Wales, Northumberland, Cumbria, Devon, Kent and Sussex suggests that some areas, at least, continued to be poorly enumerated.

The same tables as for the 1891 Census were created to test the reliability of the non-blank responses (Tables 7-9).

Status	*N*	*%*
Employer	587332	5.08
Own Account	1213611	10.49
Worker	9763192	84.43

Table 7. Employment Status of over 15s, 1901 (I-CeM).

As noted above in the discussion of the 1891 Census, interpreting change over time is difficult due to the uncertainty regarding over-counting of employers in the 1891 Census. The decline in the number and proportion of employers from 1891 probably represents a combination of more accurate data and economic change. Judging the balance between these two factors is difficult and is discussed more fully in a Working Paper on long-term trends

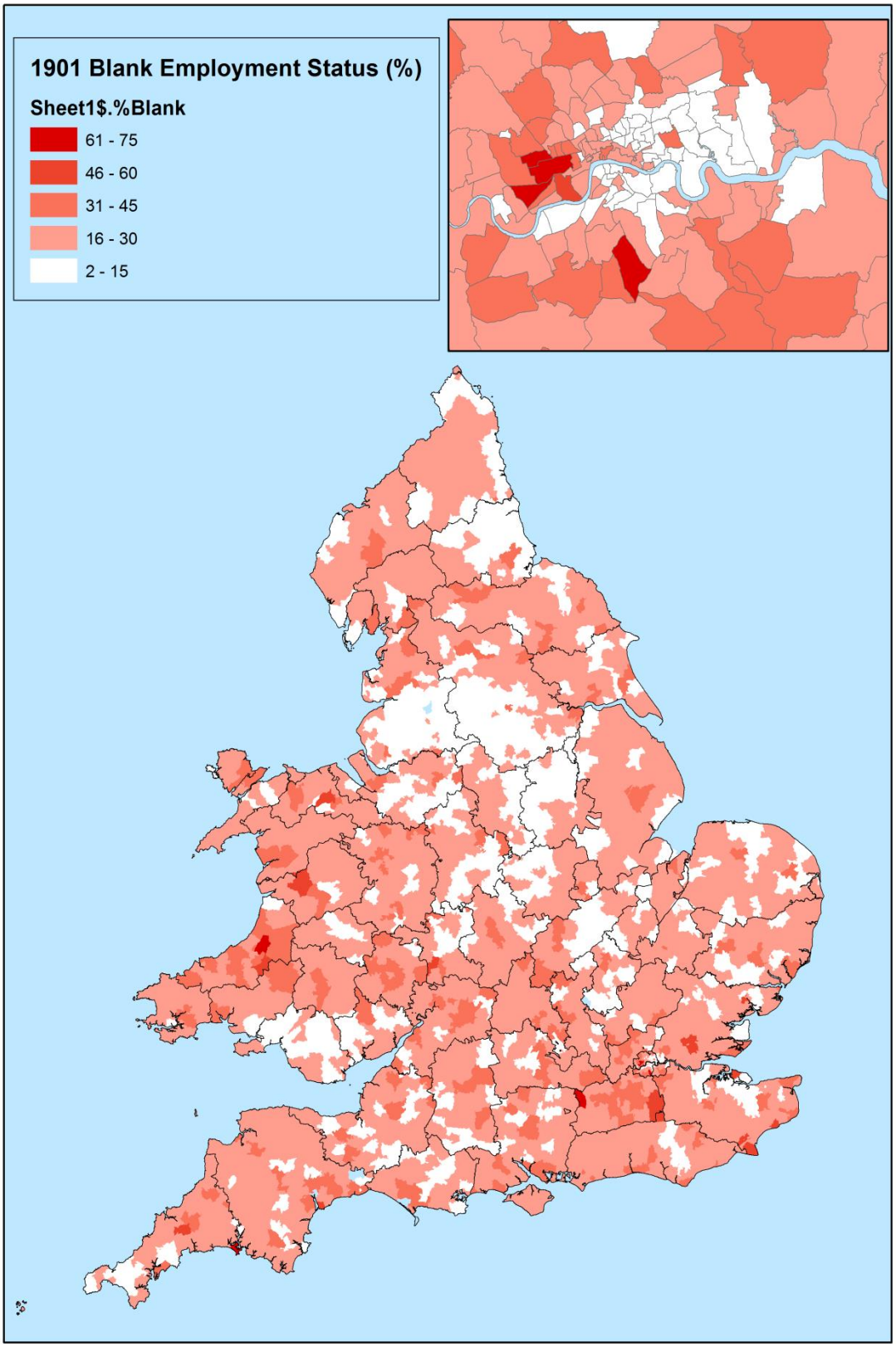

Figure 2. Blank employment status answers, percentage by RSD, 1901.
Note: The following occupation descriptors removed: '-', '*servant*', '*service*', '*retire*', '*pension*', '*own mean*', '*student*', '*scholar*', 'school', 'at school', '*(domestic)*'.

	I-CeM	*1901 Report*	*Difference*
All	8,442,660	8,828,023	-385,365
Employers	381,673	386,986	-5,313
Own Account	912,262	902,018	+10,244

Table 8. Employment status of over 10s, 1901 (I-CeM).

Note: The I-CeM employer and own account counts are made of those occupations which were tabulated by employment status in the 1901 Census, and excludes professionals, farmers and so on.

	I-CeM	*1901 Report*	*Difference*
Male	12,025,192	11,973,361	+51,831
Female	13,094,015	13,196,099	-102,084
Unknown	194,253	n/a	

Table 9. Gender of over 10s, 1901 (I-CeM).

3.3 The 1911 Census.

The 1911 Census requires more pre-cleaning than the 1891 or 1901 Censuses because the information derives from the original household returns without enumerator intervention, which results in the I-CeM employment status data being more varied and untidier. In addition the nature of the transcription varies. In the two other Censuses the employment status variable has been generated from a clean 'Employ' field provided by FindMyPast. However, that field in 1911 has a great deal of extra information which has been mistakenly transcribed. This added information takes two forms. First, there are marks added by the Census clerks when they were allocating individuals to particular employment status categories.[19] Thus, all employers who worked at home were coded '3' and this number was usually written on the householder schedule. This number has often been transcribed by FindMyPast meaning that the 'Employ' field used by I-CeM to code employment status has many entries which are similar to 'Employer3' or 'Own Account6'. Secondly, other extraneous words have been transcribed in the 'Employ' field. Some of these are the result of

[19] TNA, RG 27/8, 'Instructions to Clerks Employed on Revision of Schedules' (1911), xi.

incorrectly completed schedule (e.g. 'Home' being written in the 'Employ' column rather than the 'At Home' column) and others are misspellings (e.g. 'On Account') or terms in languages other than English (e.g. 'Gweithwraig', which is Welsh for 'Worker'). I-CeM dealt with some of these variations, but not all. This means that more than 6.8 million people were given the employment code '9' which indicates an illegible or unknown employment code. The miscoding has been corrected in the database extraction by using the clerk's coding instructions to identify the first type of mistake. This has allowed the majority to be corrected. Those that remained were re-coded by hand on the basis of the contents of their 'Employ' and 'Occ' fields.

As in the previous two censuses, a significant number of people did not answer the employment status question: 22,795,096 people out of a total population of 36,352,548. 14,097,282 of these individuals reported no occupation meaning that 8,697,814 individuals who completed the occupation question also gave no employment status. Once again, many of these did not give economically active occupations; many were either students or living on their own means. Removing these leaves 3.7 million people with genuine occupations and no employment status, 22 per cent of the 16.8 economically active million people. It is unclear why the quality of the answers to the employment status question declined in 1911. Although the 1911 I-CeM database derives from the householder schedule, not the Census Enumerators Books which form the basis for all other years, this should not affect the extent to which questions were left blank. The instructions to enumerators provided by the GRO instruct enumerators to ensure schedules were 'correctly and completely filled up'.[20] However, the elimination of enumerators' book and their replacement with summary books may have meant the enumerators paid less attention to the householders' returns and consequently overlooked missing information more regularly than they had done in 1901.[21]

As in the previous two censuses the age profile and gender of those giving no response to the employment by age distribution compared to employers, workers or the self-employed shows few significant relationships (Table 10). This follows the same pairwise Spearman rank correlation tests as in Tables 1 and 6. In this case the blanks are again randomly distributed compared to age categories for the distribution of employers and own account at $p=0.05$, but between blanks and female workers the null hypothesis cannot be accepted. At the lower

[20] TNA, RG27/8 Census of England and Wales, Instructions to the enumerator, 1911, vii.
[21] TNA, RG27/8 Census of England and Wales, Instruction to registrars, 1911, 14.

significance level of p=0.1 blank males workers and all occupied are not significantly different. Whilst the general results indicate that there is no significant difference between employer and own account proportions by age and gender at p=0.05, and hence removal of blank employment status individuals should not bias interpretation of entrepreneurship, the results indicate that as in 1891 and 1901 there is a tendency to use blanks instead of indicating workers, and for more females to have blanks. In this case it was the householders themselves who definitely introduced this feature, whereas in 1891 and 1901 we cannot be sure whether it was the householders or the enumerators.

Age	*Male*				*Female*			
categories	*Emp*	*OA*	*worker*	*blank*	*Emp*	*OA*	*worker*	*blank*
15-19	7427	9932	1263760	208815	4614	15924	724699	238248
20-29	64723	91718	2359506	350981	11647	70855	961057	489627
30-39	159014	168442	1982900	275248	16078	79874	378699	367739
40-49	169846	168036	1388800	227165	20641	83430	222192	328456
50-59	130686	136187	873896	177648	19765	69863	127178	271679
60+	100188	127053	526025	260364	21359	62194	74144	391946
Spearman	-0.03	0.09	0.66	0.66	0.26	0.31	1.0*	0.69

Table 10. Non-response (blanks) to employment status question 1911 compared to age for employers, own account, workers and all occupied; 2-way ranked Spearman correlation significance tests (* indicates significant at p=0.05).

The geographical distribution of blank answers to the employment question (Figure 3) was different from 1901. Although the distribution was more widespread, the range of values was smaller reflecting that there were fewer locations in which a lack of response was the norm even if the problem of non-response was more widespread. As noted above, this most likely reflected the fact that the data available is from household schedules rather than CEBs and, consequently, did not benefit from enumerator intervention and correction.

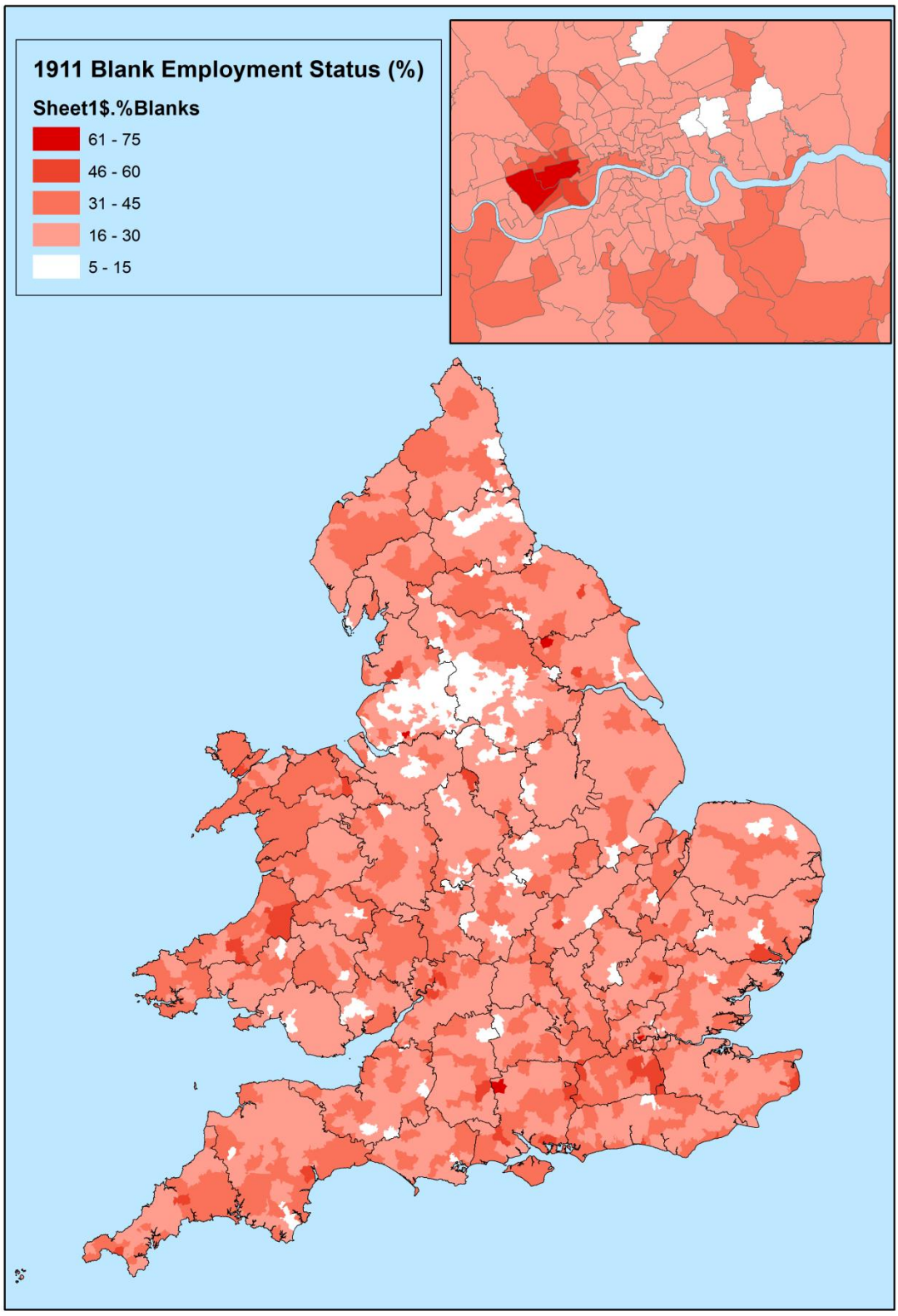

Figure 3. Blank employment status answers, percentage by RSD, 1911.

Note: The following occupation descriptors removed: '-', '*servant*', '*service*', '*retire*', '*pension*', '*own mean*', '*student*', '*scholar*', 'school', 'at school', '*(domestic)*'.

The same tables as for the 1891 and 1901 were created to test the reliability of the status responses in terms of proportions and number counts (Tables 11-13).

Status	*N*	*%*
Employer	738,322	5.63
Own Account	1,113,839	8.5
Worker	11,259,1258	85.91

Table 11. Employment Status of over 15s, 1911 (I-CeM).

These figures are similar to those for the 1901 Census. The decline in the proportion and absolute number of own account individuals and the increase in the number of employers is interesting; it is possible that this is due to increasing business concentration by size, with the self-employed being squeezed out by larger employers. This possibility is examined in detail in later Working Papers.

	I-CeM	*1911 Report*	*Difference*
All	9,844,655	10,114,162	-269,507
Employers	625,513	663,970	-38,457
Own Account	895,435	829,959	+65,476

Table 12. Employment status of over 10s, 1911 (I-CeM).

Note: The I-CeM employer and own account counts are made of those occupations which were tabulated by employment status in the 1911 Census, and excludes professionals and so on.

	I-CeM	*1911 Report*	*Difference*
Male	13,663,313	13,602,197	+61,116
Female	14,822,951	14,857,113	-34,162
Unknown	4,124	n/a	

Table 13. Gender of over 10s, 1911 (I-CeM).

3.4 Missing employment status by occupational category

It is also important to assess if there was systematic bias to the non-responses to the employment status question by occupational category. Was an enumerator or householder more likely to record a blank for some occupations rather than others? This possibility is tested in Table 14 for the three censuses, again using Spearman to test the correlation between two-way comparisons of ranked frequency of occupations for blanks against each employment status category and against all occupied. The null hypothesis is that there was a random ranked frequency relation between blanks and the occupational categories for each employment status. The occupational categories are an aggregation to 50 categories into industry 'sectors' (50ID) discussed, as defined in Working Paper 5. For males, the only significant relation was between blanks and worker categories, but this was marginal and was not significant at p=0.02. However, for females the blanks do show some significant frequency relationships with occupational categories for employers and own account statuses. For 1901, blanks were more likely to have a similar ranked frequency for all female employment statuses (except for the total of all occupied); and in 1911 blanks were more likely to be recorded for female employers and workers. Two of these categories were significant at p=0.01 as well as p=0.5. It appears that in 1901 for females blanks cannot be assumed to be either random by occupation and status, nor be mainly attributed to workers. Rather enumerators were capable of ascribing no occupation to all employment statuses of women. As in earlier comparisons it is clear that there was a systematic under-recording of female occupations, and this appears to have become worse in 1901 and 1911 compared to 1891. This appears to interrelate with household status (see below). Further statistical tests on the blanks are reported in subsequent publications.

Census	*Male*				*Female*			
	Emp	*OA*	*worker*	*total*	*Emp*	*OA*	*worker*	*total*
1891	0.144	0.018	0.284*	0.020	0.168	0.069	0.109	0.258
1901	0.132	0.156	0.032	0.128	0.298*	0.292*	0.508**	0.260
1911	0.227	0.137	0.192	0.013	0.457**	0.121	0.326*	0.207

Table 14. Non-response (blanks) to employment status question 1891-1911 by 50ID occupational status for employers, own account, workers and all occupied; 2-way ranked Spearman correlation significance tests (* indicates significant at p=0.05; ** at 0.01).

3.5 Missing employment status by relationship

The discussion above suggests that relationship to head was likely to be important to how employment status was recorded. Generally individuals, when recorded in the census by another person as head, were less likely to have their own employment status given, especially if they were female. This may reflect biases by both heads and enumerators. Whilst the census instructions themselves emphasised the importance of responses from heads but employment status of others within households was somewhat neglected. An initial exploration of this issue is developed here. It uses the 'rela' code given in I-CeM to identify individuals by the relationship in the household to the household head.[22] This is a preliminary examination pending more detailed re-coding of the rela to clean some of its mis-codings and the development of a new code (Erela) to better reflect the relationships between the entrepreneurs and others in the household (see subsequent working paper). In addition, because of the complexity and number of rela codes aggregation into broader categories is needed for statistical analysis. However, because the main patterns are so clear, for this discussion we concentrate on the most frequent 11 categories which account for most rela responses.

Relationships within households for those with blank responses and full responses to the employment question are shown in Tables 15 and 16. For males (Table 15) the most frequent categories of relationship for blank non-responses was for sons, the head himself, and grandsons, then boarders and lodgers, then other family members, servants and visitors. A similar ranking applies to all years, with some adjustments at the lower levels; but for all years the two categories of head and sons were the two most frequent blank relationships, accounting for about 80% or more of those with no status given, with sons accounting for 67-75% of blank responses. For sons, step sons and grandsons the blanks range 72-80%. The ranked pattern of blanks is reversed between these first two categories for all other employment statuses, where the head was the most common relationship given, followed by son, which demonstrates how well heads are generally recorded. The other main categories of relationship by status were similar to those of blanks in general pattern, but rank frequency differs and adjusts somewhat between years. Workers were not more similar to blanks in

[22] It should be noted that there are deficiencies in 'rela' codes due to the complexity of household structures and the algorithmic method used to allocate relationships which leads to some mis-coding by gender and status; the results quoted below are indicative rather than definitive so that all numbers and proportions must be treated as approximations.

contrast to the age and occupational comparisons shown above. It is clear that the vast majority of male blanks were for immediate family in the household.

Rela	*1891 blank*	*1891 Emp*	*1891 OA*	*1891 worker*	*1901 blank*	*1901 Emp*	*1901 OA*	*1901 worker*	*1911 blank*	*1911 Emp*	*1911 OA*	*1911 worker*
Son	75.0	4.7	9.3	29.7	72.2	5.5	8.8	29.4	67.0	6.4	9.3	28.6
Head male	8.8	89.7	79.1	49.7	10.9	89.3	82.2	50.5	11.8	87.1	80.7	52.4
Grandson	3.1	0.4	0.2	0.6	3.0	0.04	0.1	0.6	2.8	0.1	0.1	0.5
Boarders	1.9	1.0	2.4	5.2	3.0	1.6	3.2	8.5	3.1	2.2	3.4	7.5
Lodgers	1.1	0.6	2.9	4.5	0.4	0.2	0.8	1.4	0.5	0.3	1.0	1.2
Step son	1.1	0.1	0.3	1.1	1.2	0.1	0.3	1.1	1.8	0.3	0.4	1.3
Nephew	1.1	0.1	0.3	0.8	1.0	0.2	0.3	0.8	1.0	0.2	0.3	0.8
No rel given	1.0	0.1	0.2	0.4	1.1	0.1	0.1	0.3	4.5	0.1	0.3	0.5
Servants	0.9	0.1	0.1	2.1	0.9	0.1	0.03	1.8	0.7	0.1	0.03	1.4
Visitors	0.8	0.8	2.9	4.6	0.9	1.0	1.0	0.8	0.8	1.2	1.0	0.7
Total (000s)	5917	614	552	6916	6572	520	739	7774	7519	639	714	8794

Table 15. Response to employment status question by rela code for males (% of each status for that rela code) 1891-1911; ranked in order of 1891 blanks for the eleven most frequent relationships within the household.

For females (Table 16) the most frequent categories of relationship within households for non-responses to employment status in 1891 were daughters, wives, servants, then head herself, then visitors, then other family, boarders and lodgers. The four categories of daughters, wives, servants and female heads accounted for over 86% of blank occupations for each year. For daughters, step daughters and granddaughters the blanks range 42-46%. As for males, other employment statuses differ in rank frequencies among these categories, but the top three remain the main relationships; only servants drop down greatly in frequency. Workers were more similar to blanks in all years.

Rela	*1891 blank*	*1891 Emp*	*1891 OA*	*1891 worker*	*1901 blank*	*1901 Emp*	*1901 OA*	*1901 worker*	*1911 blank*	*1911 Emp*	*1911 OA*	*1911 worker*
Daughter	42.8	10.6	24.2	43.3	40.9	11.8	25.2	49.2	38.2	15.1	23.9	51.0
Wife	33.4	16.7	19.4	12.6	34.6	14.9	18.4	9.4	35.9	16.6	21.0	10.3
Servants	5.8	1.4	0.4	14.2	6.0	2.5	0.4	10.6	5.3	1.9	0.5	9.4
Head fem	5.4	61.9	39.2	9.7	6.0	60.1	41.2	9.4	6.5	53.8	38.3	7.7
Granddau	1.9	0.1	0.5	1.1	1.6	0.1	0.4	1.0	1.6	0.2	0.4	0.9
Visitor	1.4	1.0	1.8	1.3	1.4	1.1	1.6	1.6	1.3	1.9	1.8	1.4
Step dau	1.3	0.6	1.8	2.5	1.4	0.9	1.6	2.7	1.7	1.1	1.6	3.0
Niece	1.1	0.4	1.1	1.8	1.1	0.5	1.0	1.9	1.0	0.6	0.9	1.9
Boarder	1.1	0.8	1.8	3.4	1.5	1.5	2.1	5.2	1.8	2.3	2.4	5.1
Sister	1.1	3.0	3.6	2.6	1.2	3.5	4.1	2.9	1.2	3.7	4.7	3.1
Mother	0.7	0.3	0.7	0.3	0.6	0.2	0.4	0.2	0.8	0.2	0.4	0.2
Lodger	0.6	0.8	1.8	2.2	0.2	0.1	0.5	0.6	0.2	0.3	0.5	0.5
Total (000s)	12057	104	398	2352	13750	64	469	2404	15267	101	405	2904

Table 16. Response to employment status question by rela code for females (% of each status for that rela code) 1891-1911; ranked in order of 1891 blanks for the eleven most frequent relationships within the household

Overall, relationships within households are an important element accounting for higher levels of blank responses. Of these the number of servants that have blanks is less important since all are presumably of worker status. However, the pattern of blanks for male relatives especially sons, and for lodgers, boarders and visitors will affect the ability to recognise entrepreneurs. For females, as the number with blanks is a far larger than for males, there is a higher tendency to under-record employment status of female relatives, although the proportions were generally a little lower than for males and many were under 15. It is clear overall that relationship within the household was important to blank recording and this needs to be controlled for in analysis. Further statistical tests on the blanks for household relationships and how they are managed in analysis, are reported in subsequent publications.

3.6 Missing occupations

The occupational descriptor string is the key evidence base used for information on entrepreneurs and business proprietors used in the ESRC project. Unfortunately there were a number of blank entries of occupation descriptors in the census, just as there were for employer status. It is important to know if these non-responses have a systematic pattern that could bias the interpretations of the data. As with employment status, the blanks could result from enumerator and/or household omissions. Tables 17 to 19 show, respectively for 1891, 1901 and 1911, the blank responses to the occupation question by age, but where employment status is given.

Generally, the blank occupational categories are highest for those under 15, which is accounted for by the assumption by enumerators or householders that these have no occupation, but they have not written 'scholar' or similar. This accounts for 230,000 – 480,000 people across censuses, which is about 90% of all blanks for males and about 40% for females. For females, as well as the scholar category, it is clear that enumerators of householders assumed many had no employment other than in the home. This is generally highest for the 25-45 age groups; i.e. wives, and sons or daughters living at home. Probably after this age many of the single females who had been living at home with 'no occupation' became the householder themselves resulting in an occupation being more likely to be recorded, such as 'own means'.

Spearman ranked correlation tests by age between two-way comparisons of blank occupations against each employment status category and against all occupied for each gender suggests some interesting patterns. In each case the first row, for those under 15, is excluded. For male employers and own account the blank occupations were uncorrelated with age for all years. Male workers and the total of all categories were correlated with age in 1911, indicating that occupational categories were more likely to be under-recorded for younger aged men, with a decline in under-recording with age. For females, all blank occupations for employers were uncorrelated with age, except in 1911 where it is just significant (matching the critical level at $p=0.05$). However, for all other female categories (own account, workers and total workforce) under-recording, as blanks, was significantly related to age. Under-recording generally declined with age, probably because these individuals were less likely to be family servants (or they had left home).

Age	*Male*				*Female*			
categories	*Emp*	*OA*	*worker*	*blank*	*Emp*	*OA*	*worker*	*blank*
0-14	133	1895	5955	2337593	150	2165	4243	2524819
15-25	165	1151	12253	177813	301	4404	15225	953943
25-34	306	854	4762	20724	714	4961	12255	1469333
35-44	341	684	2386	12675	766	4491	10122	1217409
45-54	268	709	1405	10197	871	4691	8125	881719
55-64	175	822	908	10905	624	4225	4828	551200
65-74	135	991	489	16026	367	3196	2082	299430
75-84	48	529	149	13392	153	1257	418	97242
85+	14	78	107	8198	34	199	153	23507
Total	1585	7729	28414	2607523	3980	29589	57451	8018302
Spearman	0.14	0.73	0.64	0.64	0.71	0.92*	0.92*	0.92*

Table 17. Non-response (blanks) to occupation question in 1891 for employers, own account, workers and status blanks; 2-way ranked Spearman correlation significance tests excluding row 1 (* indicates significant at p=0.05).

Age	*Male*				*Female*			
categories	*Emp*	*OA*	*worker*	*blank*	*Emp*	*OA*	*worker*	*blank*
0-14	41	108	2472	4636977	38	115	2219	4831240
15-25	54	142	3634	116763	100	601	7733	1179568
25-34	64	119	1234	20446	191	907	4297	1831304
35-44	67	121	723	12730	207	958	3229	1499249
45-54	55	116	426	9724	209	960	2580	1068114
55-64	44	100	280	11164	184	793	1574	686459
65-74	30	114	170	17294	108	452	669	358327
75-84	5	55	62	15613	34	191	154	125965
85+	9	57	71	5045	12	17	32	20565
Total	364	877	9010	4845756	1083	4994	22487	11600791
Spearman	0.20	0.52	0.52	0.62	0.29	0.79*	0.93*	0.93*

Table 18. Non-response (blanks) to occupation question in 1901 for employers, own account, workers and status blanks; 2-way ranked Spearman correlation significance tests excluding row 1 (* indicates significant at p=0.05).

Age	*Male*				*Female*			
categories	*Emp*	*OA*	*worker*	*blank*	*Emp*	*OA*	*worker*	*blank*
0-14	83	452	2926	3266416	82	627	3154	3395081
15-25	105	297	7285	169238	188	2034	15329	900716
25-34	174	273	2828	77254	371	2733	9343	1893632
35-44	154	261	1769	23968	561	3252	8181	1708808
45-54	151	213	1158	9111	587	3066	7176	1207669
55-64	106	202	755	8515	483	2126	3993	776450
65-74	62	165	422	14419	342	1160	1488	449760
75-84	22	59	96	11092	101	261	240	145803
85+	9	22	55	6776	30	46	54	26069
Total	866	1944	17294	3586779	2745	15305	48958	10503988
Spearman	0.04	0.18	0.81*	0.81*	0.74*	0.87*	0.86*	0.86*

Table 19. Non-response (blanks) to occupation question in 1911 for employers, own account, workers and status blanks; 2-way ranked Spearman correlation significance tests excluding row 1 (* indicates significant at p=0.05).

The overall pattern is disturbing, although to be expected from previous discussions of the census. It indicates that females were again more likely to be recorded with no occupation, and under-recording of occupations was generally related in rank frequency to age. However, for men blank occupational responses were randomly distributed by age compared to each employer category, though there was a relationship of age to blank occupations for workers and all occupied in 1911. For the target categories of *male* employers and own account the comparison indicates that non-responses to the occupation question were randomly distributed by age and status throughout the population; hence removal of those who failed to provide an occupation before subsequent statistical analysis should not bias interpretation. But for *females* a different approach is required to take account of a high frequency of under-recording of occupations in general, and of own account status for all years, and for female employers in 1911. Further statistical tests on missing occupations and alternative methods of managing non-responses are reported in subsequent publications.

3.7 Occupation coding 1891-1911

In order to ensure the occupation coding of the employers and own account individuals is as accurate as possible two checks were performed. First, the total numbers of employers and own account in each census occupation category were extracted from I-CeM and then compared to the totals given in the census reports.[23] For any category where the ratio of I-CeM total to report total was greater than 2 the most common occupation descriptors were extracted. All occupation strings in these problematic categories which either described more than 100 employers or own account or were more than 5 percent of the total number in that category were checked and, if necessary, corrected. The occupation coding in I-CeM are consistent between all censuses; thus, if an occupation is incorrectly coded in one census it will have the same occupation code in all other censuses. Therefore, all incorrect coding identified in each individual census were applied to all other censuses. Furthermore, as many of the employers' and own account individuals' occupation descriptors were shared by non-entrepreneurs, workers and those with blank employment statuses had their occupation codes changed and consequently the overall quality of the occupation coding of the entire population was improved.

Since the census reports do not provide breakdowns by employment status for all occupation categories (the professions, for example, were never broken down) it was necessary to undertake a second check. All occupation strings that described 25 or more employers or own account individuals were extracted, their coding checked, and corrected if necessary. Again, the corrections from one census were applied to all other censuses.

Tables 20 and 21 detail the number of strings checked, the number of strings that were changed and the total number of individuals affected.

[23] *Census of England and Wales, 1891, Ages, condition as to marriage, occupations, birth-places and infirmities, Vol. III*, *PP*, 1893-4, (CVI), x-xxv; *Census of England and Wales, 1901, Summary Tables. Area, houses and population; also population classified by ages, condition as to marriage, occupations, birthplaces and infirmities*, *PP*, 1903 (LXXXIV), 186-201; *Census of England and Wales, 1911, Vol. X, occupations and industries*, *PP*, 1913 (LXXVIII), 12-25.

Year	*Total strings*	*Total E/OA*	*N of common strings*	*E/OA with common strings*	*% of total E/OA checked*	*No. of strings changed*
1891	139,208	1,620,266	3,071	1,383,606	85.39	141
1901	184,628	1,796,599	3,720	1,464,394	81.51	141
1911	283,257	1,832,671	4,049	1,383,207	75.47	170

Table 20. Checking common occupation strings, 1891-1911

Year	*Total no. of strings changed*	*No. of people with corrected occode*	*% of population with an occupation with corrected occode*
1891	478	308,417	1.68
1901	481	258,965	1.62
1911	432	309,791	1.40

Table 21. Total strings changed, 1891-1911

These checking processes ensure that at least three quarters of the entrepreneurs in each census, and all of those with strings common enough to be shared among 25 people, have the correct occupation code. In each case the actual proportion with correct occupation coding will be higher than given in Table 21 as many of those strings that described fewer than 25 employers or own account will already be correctly coded. In addition all employers with portfolios of activity were checked by hand, which covers about 10% of all employers. However, it is impractical to check strings that described very few people as the vast majority of strings described just one individual, as shown in Table 22. In each census, the long tails of strings which describe fewer than ten individuals are very difficult to check or correct in any systematic manner.

No. of people described by string	*No. of strings*			*% of total strings*		
	1891	*1901*	*1911*	*1891*	*1901*	*1911*
1	109,215	141,453	228,090	77.83	76.62	80.52
2-10	25,301	35,915	46,921	18.06	19.45	16.56
11-20	2,293	2,981	3,536	1.63	1.61	1.25
21-30	914	1,124	1,281	0.65	0.61	0.45
31-40	425	654	739	0.30	0.35	0.26
41-50	321	388	416	0.22	0.21	0.15
51-100	756	857	970	0.54	0.46	0.34
101-200	483	547	557	0.34	0.30	0.20
201-500	304	359	378	0.22	0.19	0.13
501-1000	137	156	174	0.10	0.84	0.06
1000+	173	193	195	0.12	0.10	0.07

Table 22. Entrepreneur string frequency, 1891.

3.8 Cleaning 1891-1911

The same processes have been applied to each census to clean the data and extract the employers, own account and workers into separate databases for subsequent analysis. The following steps were followed:

1) All individuals with either no reported occupation or no answer to the employment status question were removed.
2) All individuals under the age of fifteen were removed.
3) All boarders, visitor and lodgers. These are tagged so that they can be removed in some analyses when not required.
4) All individuals in institutions or aboard ship. These are tagged so that they can be removed in some analyses when not required.

5) All non-economically active individuals were removed. They were identified using wild-card searches: *retire*, *own mean*, *pension*, *former*, *unemploy*, *student*, *scholar*, *pupil*.
6) All individuals in the occupation codes (I-CeM occodes) listed in Table 23 were removed. These codes cover individuals who were not economically active and also those whose occupation are excluded in subsequent analysis of UK entrepreneurs (such as foreign diplomats).

Occode	*Occupation Category*
772	RETIRED (NOT ARMY OR NAVY)
773	ARMY PENSIONERS
774	NAVY PENSIONERS
775	PENSIONERS, SUPERANNUATED
776	RECEIVING OLD AGE PENSION (OCCUPATION OR FORMER OCCUPATION NOT STATED)
778	PRIVATE MEANS
779	THEOLOGICAL STUDENTS
780	LAW STUDENTS
781	MEDICAL STUDENTS
782	LITERARY STUDENTS
783	ART STUDENTS
784	SCIENTIFIC STUDENTS
785	OTHER STUDENTS
786	AGRICULTURAL STUDENTS
787	SCHOLARS ETC
788	WIVES AND OTHERS ENGAGED IN (OWN) HOUSEHOLD DUTIES
790	WIDOWS (OF NO SPECIFIED OCCUPATION)
791	CHILDREN RELATIVES VISITORS AT HOME
793	FOREIGN DIPLOMATS
794	NO SPECIFIED OCCUPATION - RECEIVING INCOME, SUPPORT
795	NO SPECIFIED OCCUPATION - CHILDREN
796	PRISONERS, REFORM SCHOOL INMATES ETC
797	NO SPECIFIED OCCUPATION - VAGRANTS, UNEMPLOYED

Table 23. I-CeM Occodes removed from entrepreneur database.

Following these six steps, the employers and own account individuals in each census were further screened by the following steps.

1) All employers and own account individuals with any of the following terms within their occupational descriptor were removed on the basis that their occupation meant they were not legitimate employers or own account. They were identified by wild-card searches: *apprentice*, *journeyman*, *servant*, *labourer*, *clerk*, *assistant*, *attendant*, *mechanic*, *artisan*, *machinist*.
2) All employers and own account individuals in any of the I-CeM occodes listed in Table 25 were also removed. Again, these occodes are those that are considered to contain incorrectly coded employers or own account individuals.

In most cases, the occodes cleaned in Table 23 reflect decisions to focus in the entrepreneurship analysis on those individuals who are generally bearing the risk of their enterprise themselves. In line with modern analyses, therefore, those individuals were screened who were de facto employees of state enterprises (such as the post office, most schools, etc.) or companies (such as company secretaries and managers) or were pursers or managers of enterprises or officers of local boards. In addition, those occupations where a small self-employment income was available as a by-product of their employment were also screened; this applies to clergy of the various churches who took personal fees for ceremonies and over 3,000 of them were returned as employers and own account in 1891.

In some other cases, occupations with legitimate answers to the employment status question may have been coded to the wrong I-CeM occupational code. This is confirmed in another Working Paper which compares I-CeM, CEB and local directory entries in a series of case study areas. The screened categories therefore contain very small numbers of genuine employers or own account. However, for some categories, such as nurses, mid-wives, and jobbing gardeners, additional analysis is included in some subsequent research.

The list in Table 24 errs on the side of caution on the basis of a judgement that it is preferable to exclude a few genuine employers and own account individuals than to include very large numbers of false positives.

Table 24. Non-Employer or Own Account I-CeM Occodes.

Occode	*Occupation Category*
1	POST OFFICE - TELEGRAPHISTS, TELEPHONE OPERATORS
2	OTHER POST OFFICE OFFICERS AND CLERKS
3	POSTMEN
4	POST OFFICE MESSENGERS, ETC
5	MPs, MINISTERS OF THE CROWN & PEERS
6	OTHER CIVIL SERVICE OFFICERS AND CLERKS
7	PRISON OFFICERS
8	SENIOR OFFIALS AND OTHERS IN EAST INDIA SERVICE
9	OTHER CIVIL SERVICE MESSENGERS, ETC. (INC PORTERS),
10	PRISON SERVICE MESSENGERS, ETC.,
11	POLICE
12	POOR LAW SERVICE
13	MUNICIPAL, PARISH, AND OTHER LOCAL OR COUNTY OFFICERS
14	MUNICIPAL, PARISH, AND LOCAL OR COUNTY SHERIFFS AND CLERKS
15	MUNICIPAL, PARISH, AND OTHER LOCAL OR COUNTY WORKERS
17	ARMY OFFICERS (EFFECTIVE)
18	MILITIA & YEOMANRY OFFICERS (EFFECTIVE)
19	ARMY OFFICERS (RETIRED)
20	SOLDIERS AND NON-COMMISSIONED OFFICERS
21	MILITIA & YEOMANRY MEN
22	OFFICERS OF THE NAVY (EFFECTIVE)
24	MEN OF THE NAVY (INC COASTGUARDS)
25	OFFICERS OF THE MARINES (EFFECTIVE)
27	MEN OF THE MARINES
28	CLERGYMAN OF THE ESTABLISHED CHURCH (CHURCH OF ENGLAND IN ENGLAND AND WALES, CHURCH OF SCOTLAND IN SCOTLAND)
29	ROMAN CATHOLIC PRIEST
30	MINISTER, PRIEST, OF UNITED FREE CHURCH
31	MINISTER, PRIEST, OF EPISCOPALIAN CHURCH IN SCOTLAND
32	MINISTER, PRIEST, OF UNITED PRESBYTERIAN CHURCH
33	MINISTER, PRIEST, OF OTHER RELIGIOUS BODIES
34	ITINERANT PREACHER, SCRIPTURE READER, MISSION WORKER

35 MONKS

36 NUN, SISTER OF CHARITY

37 CHURCH, CHAPEL, CEMETERY--OFFICER, ETC.

40 LAW CLERK

45 MIDWIVES

46 NURSES (MEDICAL NOT DOMESTIC) INCLUDING POOR LAW NURSES

47 HOSPITAL SICK NURSES

48 MEDICAL ASSISTANTS

50 HOSPITAL WORKERS GENERAL - DEFAULT

51 HOSPITAL MEDICAL SERVICE

54 SCHOOL SERVICE (PUPIL TEACHERS)

55 OTHERS CONNECTED WITH EDUCATION - SCHOOL SERVICE GENERAL, ATTENDANTS, CLERKS, ETC

78 ARCHITECTS ASSISTANTS

82 DOMESTIC INDOOR SERVANTS IN HOTELS, LODGING HOUSES AND EATING HOUSES

83 OTHER DOMESTIC INDOOR SERVANTS -- GOVERNESSES

84 OTHER DOMESTIC INDOOR SERVANTS -- UNDEFINED

85 DOMESTIC--COACHMAN

86 DOMESTIC--MOTOR CAR DRIVER, MOTOR CAR ATTENDANT

87 DOMESTIC GARDENER

88 GAMEKEEPER

89 ARMY COLLEGE, CLUB--SERVICE

90 OTHER COLLEGE, CLUB--SERVICE

91 PRISON OFFICERS (REFORMATORY SCHOOL)

93 HOSPITAL, INSTITUTION (NOT POOR LAW), AND BENEVOLENT SOCIETY--SERVICE (NOT SICK NURSES, MEDICAL MEN, OR SCHOOLMASTERS)

94 PARK, LODGE, GATE, ETC.--KEEPER (NOT GOVERNMENT)

95 CHURCH CHAPEL CEMETERY CARETAKERS KEEPERS

97 CARETAKER, OFFICE KEEPERS (NOT GOVERNMENT) - SCHOOL CLEANERS, CARETAKERS

99 COOKS (COLLEGE: NON-RESIDENT)

100 COOKS (BOARDING LODGING HOUSE: NON-RESIDENT)

101 COOKS (DOMESTIC: NON-RESIDENT)

102 DAY GIRLS, DAY SERVANTS (OTHER INDOOR)

103 DAY GIRLS, DAY SERVANTS (HOTELS LODGING EATING HOUSES)

107 GOVERNESSES (DOMESTIC)(NON-RESIDENT)

108 SERVANTS - UNDEFINED (NON-RESIDENT)

109 OTHERS IN SERVICE

119 COMMERCIAL OR BUSINESS CLERKS

125 RAILWAY OFFICIAL - INSPECTORS, SUPERVISORS, CLERK

126 RAILWAY TICKET-EXAMINER, COLLECTOR, CHECKER

127 RAILWAY ENGINE--DRIVER STOKER CLEANER

128 RAILWAY GUARD

129 SIGNALMAN

130 POINTSMAN, LEVEL CROSSING MAN, GATEKEEPERS

131 PLATELAYER, GANGER, PACKER

132 RAILWAY LABOURER (NOT RAILWAY CONTRACTOR'S LABOURER)

133 RAILWAY PORTER

134 OTHER RAILWAY SERVANTS

150 OMNIBUS CONDUCTORS

155 OTHERS ON ROADS -- TOLL COLLECTORS

159 MERCHANT SERVICE: SEAMAN--COOKS, STEWARDS, AND OTHERS (SUBSIDIARY SERVICE)

163 CANAL AND INLAND NAVIGATION SERVICE (ON SHORE)

166 STEVEDORES, WHARF AND DOCK LABOURERS

170 COALHEAVER; COAL--PORTER, LABOURER

172 TELEGRAPH, TELEPHONE--SERVICE (NOT GOVERNMENT)

175 FARMER'S, GRAZIER'S--SON, DAUGHTER, OR OTHER RELATIVE ASSISTING IN THE WORK OF THE FARM

176 CROFTERS--SON, DAUGHTER, OR OTHER RELATIVE ASSISTING IN THE WORK OF THE FARM

178 SHEPHERD

179 AGRICULTURAL LABOURER, FARM SERVANT--DISTINGUISHED AS IN CHARGE OF CATTLE

180 AGRICULTURAL LABOURER, FARM SERVANT--DISTINGUISHED AS IN CHARGE OF HORSES

181 AGRICULTURAL LABOURER, FARM SERVANT--NOT OTHERWISE DISTINGUISHED

197 SHALE MINER

199 RAILWAY LABOURERS NAVVIES (COAL MINE)

205 FACTORY LABOURERS (UNDEFINED) COKE AND GAS

266 FITTERS, TURNERS (ENGINE AND MACHINE) LABOURERS

274 LABOURERS (UNDEFINED) IN ENGINEERING WORKS

283	TECHNICIANS (UNSPECIFIED)
354	SHIPYARD LABOURERS (UNDEFINED)
407	BUILDER'S LABOURER
410	CARPENTER'S, JOINER'S--LABOURER
413	BRICKLAYER'S LABOURER
415	MASON'S LABOURER
418	PLASTERER'S LABOURER
431	SEA WALL GROYNE LABOURERS PILE DRIVERS
432	RAILWAY LABOURERS NAVVIES (CONTRACTORS LABOURERS) DEFAULT
436	ROAD LABOURERS
608	FACTORY HANDS (TEXTILE) UNDEFINED (VARIOUS)
717	BARMEN (NOT IN SERVICE)
718	BOARDING LODGING HOUSE WAITERS (NON-RESIDENT)
719	HOTEL WAITERS
720	HOTEL PORTERS BOOTS KNIFEMEN PLATEMEN PLATE CLEANERS POLISHERS (NON-RESIDENT)
721	OTHER HOTEL SERVANTS
764	CORPORATION BOROUGH COUNCIL LABOURERS (UNDEFINED)
765	GENERAL LABOURERS
766	ENGINE DRIVERS, STOKERS, FIREMEN (NOT RAILWAY, MARINE, OR AGRICULTURAL)
769	APPRENTICES
770	FACTORY LABOURERS (UNDEFINED)
789	WIVES ASSISTING GENERALLY IN THEIR HUSBANDS OCCUPATIONS (WIFE OF …)
792	PROSTITUTES
793	FOREIGN DIPLOMATS

4. Conclusion.

This paper examines the use of the 1891-1911 censuses to identify employers and own account self-employed. The extraction of individuals from these censuses provides the basis of the database for the ESRC project. Other Working Papers examine further details of

aspects of the extraction and coding of individuals and occupations to yield the final database and variables that can be used for statistical analysis.

We find no evidence of systematic errors in the 1891 responses to the employment status question, contrary to the claims made by GRO at the time. There are few double ticks and all appear to relate to genuine double occupations. However, a key element to take forward from this paper is the need to exercise care when interpreting non-responses (blank entries) to the employment status and occupations questions. Compared to ticking the wrong boxes, these have large frequency. Blank responses to employment status have a geographical pattern reflecting different enumerator capacity which indicates the need for careful analysis, especially at the smallest level of the parish. More generally blank employment status appears to correlate most closely to worker status, especially for females. But there is a concern that some employers may be excluded because of the tendency of some enumerators or householders to largely ignore the status column and leave it blank, especially in 1901, and especially for females for all years. It is also clear that there were some occupational and relationship biases of blank responses to employment status for females. There may also be an issue of mis-attribution between employer and own account status in 1891, as suggested by GRO. For the target categories of *male* employers and own account the non-responses were randomly spread by age and status throughout the population; hence removal of those who failed to provide occupation status before subsequent statistical analysis should not bias interpretation. But for *females* there is a high frequency of under-recording of occupations in general, and of own account status for all years, and for female employers in 1911. In addition, many within-household relatives to the head do not have status information (especially, sons, daughters, wives, brothers, and sisters), and this is a major problem for females. Boarders, lodgers and visitors also account for a large proportion of bank occupations responses. Further statistical tests on missing occupations and alternative methods of managing non-responses as a potential source of bias by weighting and other methods are reported in subsequent working papers and other publications.

As noted at the outset, the population census was not a business census, with the result that the way in which the information was gathered constrains the business information that can be obtained and prevents it being totally complete. However, as is shown in this paper, the 1891-1911 censuses provide considerable potential for identifying entrepreneurs as a whole, and for

differentiating between the different categories of employers and the own account self-employed.

Acknowledgments:

This research has been supported by the ESRC under project grant ES/M010953: **Drivers of Entrepreneurship and Small Businesses.** It also draws from piloting of the research for 1881 from Leverhulme Trust grant RG66385: **The long-term evolution of Small and Medium-Sized Enterprises (SMEs).**

The database used for 1891 and 1901-11 derives from K. Schürer, E. Higgs, A.M. Reid, E.M Garrett, *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2)*, (2016) [data collection]. UK Data Service, SN: 7481, http://dx.doi.org/10.5255/UKDA-SN-7481-1; enhanced; E. Higgs, C. Jones, K. Schürer and A. Wilkinson, *Integrated Census Microdata (I-CeM) Guide*, 2nd ed. (Colchester: Department of History, University of Essex, 2015).

The data used for the 1881 pilot and in this paper derives from Schürer, Kevin and Woollard, Matthew (University of Essex) (2000) *1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version)* [computer file] UKDA, SN-4177, supported by ESRC, derived from the manuscript census enumerators' books, transcribed by Genealogical Society of Utah and Federation of Family History Societies.

The GIS boundary files for RSDs were constructed by Joe Day for the ESRC fertility project directed by Alice Reid:
http://www.geog.cam.ac.uk/research/projects/victorianfertilitydecline/publications.html

These used as a starting point the GIS parish files of Satchell, A.E.M., Kitson, P.M.K., Newton, G.H., Shaw-Taylor, L., Wrigley E.A. (2006) *1851 England and Wales census parishes, townships and places*, 2006, ESRC RES-000-23-1579, supported by Leverhulme Trust and the British Academy; Satchell, A.E.M. (2015) *England and Wales census parishes, townships and places*; which is an enhanced and corrected version of Burton, N, Westwood J., and Carter P. (2014) *GIS of the ancient parishes of England and Wales, 1500-1850*, UKDA, SN 4828; which is a GIS version of Kain, R.J.P., and Oliver, R.R. (2001) *Historic parishes of England and Wales: An electronic map of boundaries before 1850 with a gazetteer and metadata*, UKDA, SN 4348.

A special acknowledgement of thanks is made to Kevin Schürer for advice and all his help in developing improved versions of I-CeM, and to Alice Reid, Eilidh Garrett, Joe Day, Hanna

Jaadla, Xuesheng You, Leigh Shaw-Taylor and other members of the Campop I-CeM group who, with the current authors, have collectively worked on the new versions of I-CeM.

References.

Booth, Charles (1892-1903), *Life and Labour of the People of London*, 17 vols., London: Macmillan.

Higgs, E. (1996), 'The Statistical Big Bang of 1911: Ideology, Technological Innovation and the Production of Medical Statistics', *Social History of Medicine*, 9/3 (1996), 409-26.

Higgs, E., Jones, C., Schürer, K. and Wilkinson, A. (2015), *Integrated Census Microdata (I-CeM) Guide*, 2nd ed., Colchester: Department of History, University of Essex.

Schürer, K. (1991), 'The 1891 Census and Local Population Studies', *Local Population Studies*, 47, 16-29.

Schürer, K., Higgs, E., Reid, A.M. and Garrett, E.M. (2016), *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2)*, [data collection]. UK Data Service, SN: 7481, http://dx.doi.org/10.5255/UKDA-SN-7481-1

Other Working Papers:

Working paper series: ESRC project ES/M010953: *'Drivers of Entrepreneurship and Small Business'*, University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

WP 1: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Drivers of Entrepreneurship and Small Businesses: Project overview and database design.*

WP 2: Bennett, Robert J., Smith Harry J. and van Lieshout, Carry (2017) *Employers and the self-employed in the censuses 1851-1911: The census as a source for identifying entrepreneurs, business numbers and size distribution.*

WP 3: van Lieshout, Carry, Bennett, Robert J., Smith, Harry J. and Newton, Gill (2017) *Identifying businesses and entrepreneurs in the Censuses 1851-1881*.

WP 4: Smith, Harry J., Bennett, Robert J., and van Lieshout, Carry (2017) *Extracting entrepreneurs from the Censuses, 1891-1911*.

WP 5: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Business sectors, occupations and aggregations of census data 1851-1911*.

Full list of all current Working Papers available at:
http://www.geog.cam.ac.uk/research/projects/historyofentrepreneurship/